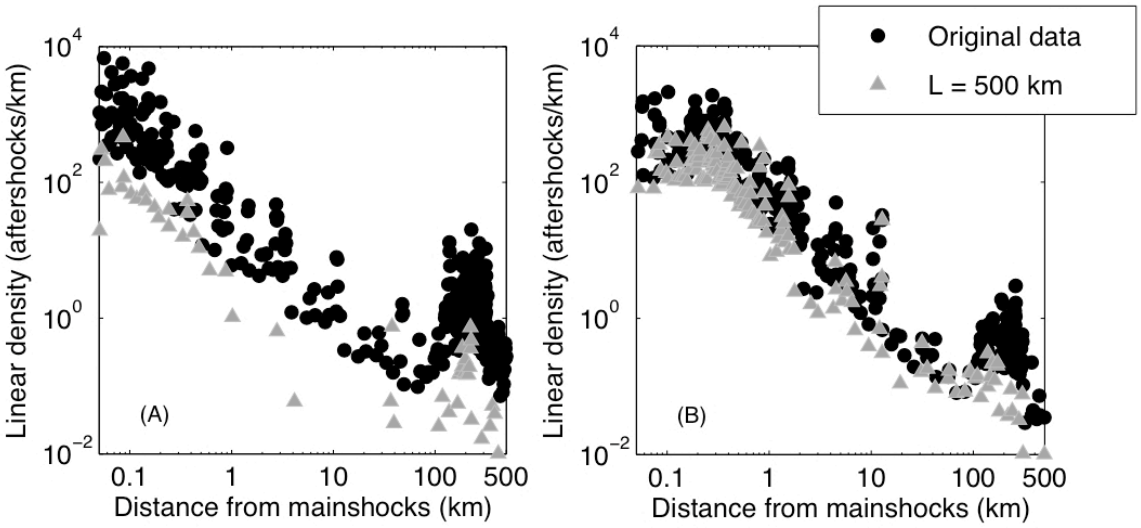


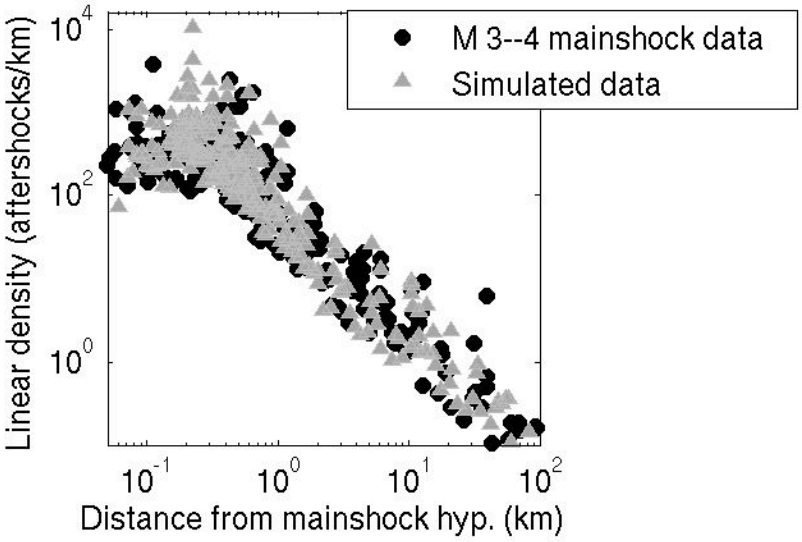
Supplementary Figure 1

Complete distance range data for (A) 2≤M<3 Southern California mainshocks and (B) 3≤M<4 Southern California mainshocks. In Figure 2 in the main text we show the 5 minute aftershocks of M 2—4 mainshocks over the range 0.2 km – 50 km. This is the range in which the data is free of artifacts and not contaminated by aftershocks of larger earthquakes or other sources of background seismicity (see Methods). Thus over this range we can accurately measure the decay of aftershock density with distance. In this Supplementary figure, for the interested reader, we provide the densities of the first five minutes of M>2 aftershocks over the fullest possible distance range, from 0 to 500 km. Black dots are data as analyzed in Fig. 2; Grey triangles are the same data with an alternative mainshock exclusion criterion of L=500 km (see Methods).

At distances less than a fault length the observed density is constant rather than following the power law discussed in this paper. This near field behavior is a result of measuring aftershock distances from the hypocenter rather than the fault plane – in the near field the point source approximation breaks down (see Supplementary Figure 2 and Figure 3 in the main text). Because mainshock fault length is important, the decay leveling is seen at longer distances for the 3≤M<4 than for the 2≤M<3 mainshocks.

At far distances there are more background earthquakes than aftershocks, and thus the aftershock decay can no longer be observed. Since the mainshock selection criteria only eliminated larger mainshocks within 100 km (see Methods), there is contamination from aftershocks of larger mainshocks beyond 100 km, as well as contamination from other sources of background seismicity. Changing to a 500 km exclusion zone for larger earthquakes decreases the number of M 2—4 mainshocks that we can use but results in a proportionally lower background seismicity rate at distance; the fraction of earthquakes occurring beyond 100 km drops by half (grey triangles). Beyond 300 km there is a precipitous drop in earthquake density because of edge effects; we begin to frequently overstep the spatial bounds of the catalog.

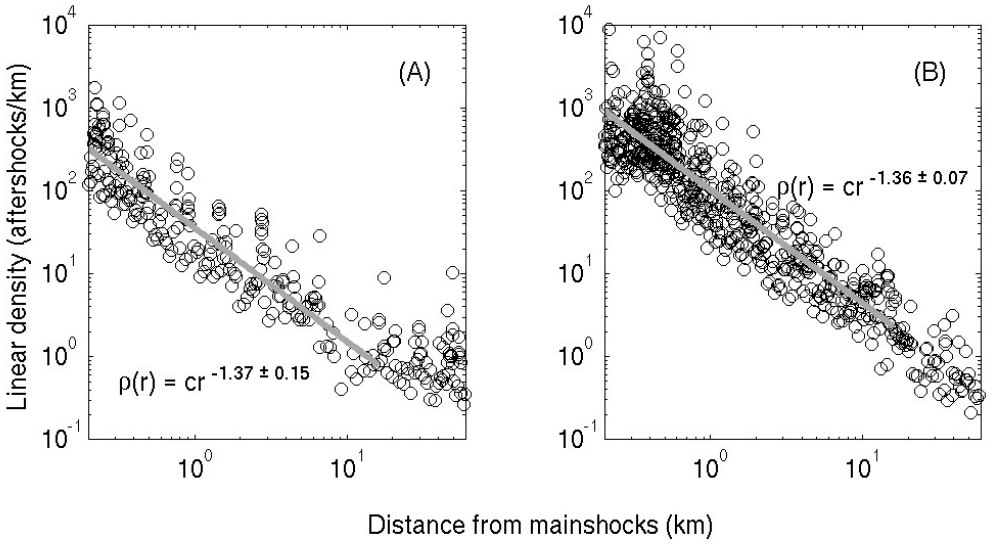
Black dots in (A) comprise 7396 mainshocks and 440 after-quakes (aftershocks plus background); gray triangles 1019 mainshocks and 50 aftershocks. Black dots in (B) comprise 2355 mainshocks and 348 after—quakes, and gray triangles 1252 mainshocks and 155 after-quakes.



Supplementary Figure 2

Demonstration via simulation that the decrease in the decay rate at short distances is due to hypocentral versus fault plane distance measurements. For the Monte Carlo simulation data (gray triangles) aftershocks are assigned a power law distribution of distances from points on planar vertical faults. The same number of mainshocks and same distribution of mainshock magnitudes is used as in the real data, and the strike of each mainshock is allowed to vary randomly. A normally distributed location error with a standard deviation of 20 m is added to each aftershock location. Distances are measured from the aftershock hypocenter to the mainshock hypocenter. We do not simulate variations in mainshock dip.

Catalog incompleteness is modeled to mimic observations of the five minute aftershock catalog data where completeness is determined by comparison with the Gutenberg-Richter relationship. In the simulation 75% of the aftershocks of M 3—3.5 mainshocks closer than one fault length are eliminated and 93% of aftershocks of M 3.5—4 mainshocks closer than one fault length are eliminated. As stated in the Methods section, in the distance range used to fit Equation 1 (0.2-50 km), fewer aftershocks are missing and the percentage of missing earthquakes is not strongly distance dependent. Simulation trials in which no earthquakes are eliminated also show the key features of this figure, i.e., a corner and decrease in the decay rate at one fault length due to the break down of the point source approximation.

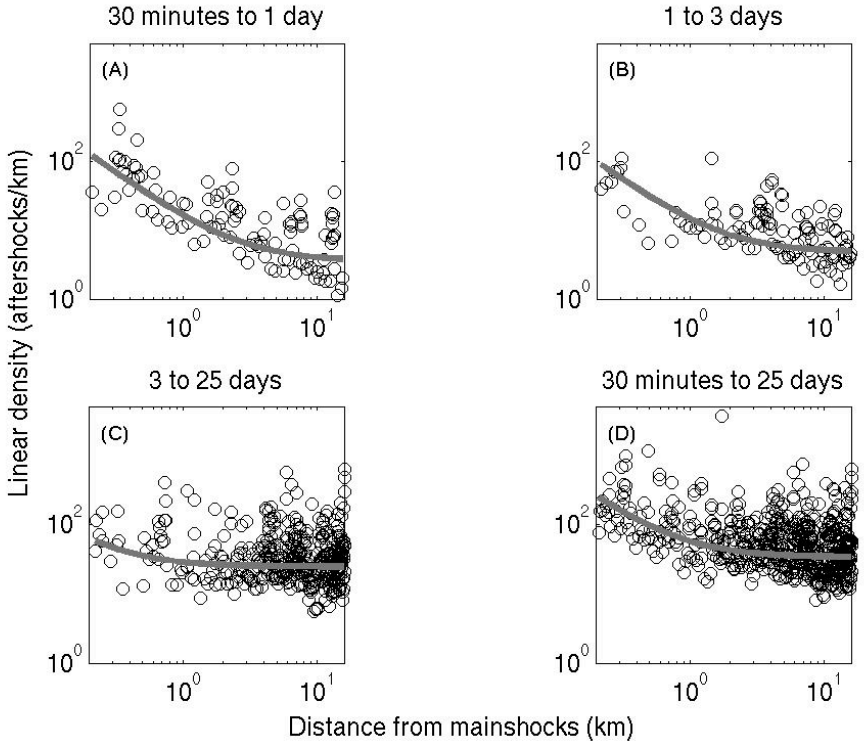


Supplementary Figure 3

Distance from the mainshock vs. linear aftershock density for 30 min of aftershocks after each mainshock. All aftershocks are M ≥ 2. The data is fit with an inverse power law (solid gray lines) over 0.2-16 km. Background interference is small for these datasets in this distance range (see below). (A) Results for 2≤ M <3 mainshocks. This plot uses 7396 mainshocks and 219 aftershocks. (B) Results for 3≤M<4 mainshocks. This plot uses 2355 mainshocks and 610 aftershocks.

There are many more mainshocks than observable aftershocks. This is because minimum aftershock magnitude is kept very close to the mainshock magnitude to ensure catalog completeness at distances used for the fit, so the large number of aftershocks that are much smaller than their mainshock are not included. The observed number of aftershocks is consistent with standard productivity relations for the small mainshock magnitudes [20, 24-25].

Background seismicity becomes apparent for aftershocks of the 2≤M<3 mainshocks at a closer distance (16 km) than it does for aftershocks of the the 3≤M<4 mainshocks (> 50 km) because many more 2≤M<3 mainshocks are used in the analysis, thus the total time represented in the plot (30 minutes per mainshock x 7396 mainshocks = 154 days) is longer than the total time used for the 3≤M<4 plot (49 days). This longer time allows more background earthquakes to accumulate. The larger quantity of background earthquakes means that the aftershock signal is obscured at shorter distances in the far field where aftershock numbers are small. The larger background is not a significant factor in the near field, where the aftershock to background earthquake ratio is high.



Supplementary Figure 4

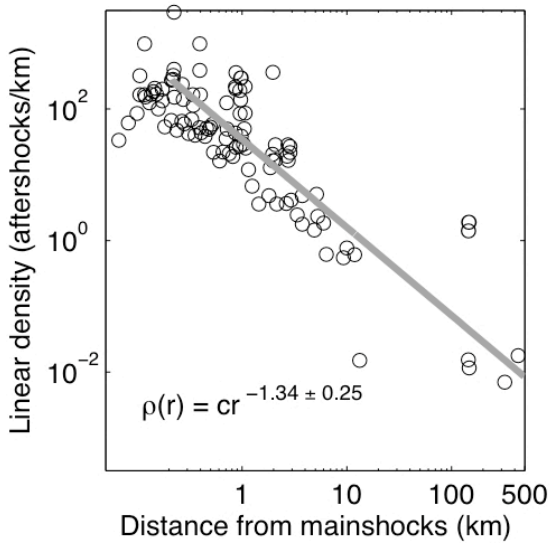
Aftershocks over longer time windows. Data is from the Shearer et al. relocated Southern California catalog; mainshocks are M 3-4, aftershocks are M>2. We study distances from 0.2 – 16 km, following Supplementary Figure 3. Because the aftershock data covers a longer time period than in Supplementary Figure 3, however, stricter rules are used with respect to making sure that mainshocks chosen are isolated from larger earthquakes, and so fewer mainshocks can be used (see Methods). There are 191 mainshocks in this data set.

The longer time spans include a larger number of background earthquakes. We fit the data with $\rho(r)=cr^{-1.4} + b$, where the fit parameters c and b are controlled by the number of aftershocks and background earthquakes, respectively. We minimize the least-square error to fit b and c. Because aftershock sequences decay rapidly we expect the percentage of aftershocks to drop with time.

(A) Data from 30 minutes to 1 day after the mainshock is best fit with 48% aftershocks, given by the gray line (c = 13.0, b = 3.5). A very small correlation coefficient of the data residuals (r = -0.004) and a Kolmogorov-Smirnoff test (at 95% and 65% confidence) confirms a good fit. There are 107 aftershocks in this plot. (B) Data from 1-3 days is best fit with 34% aftershocks, c=9.9, b=4.7; r = -0.008. 114 aftershocks used. (C) From 3-25 days, the best fit is with 4% aftershocks, c=4, b=25; r = 0.007. 452 aftershocks are used. (D) Combined 30 minutes – 25 days, c=24, b =33. 630 aftershocks are in this plot. If our previous solutions are correct, then 17% of this total should be aftershocks; the best fit gives 15%. As an additional check of consistency we do a weighted average of the daily background rate from panels (A) through (C),

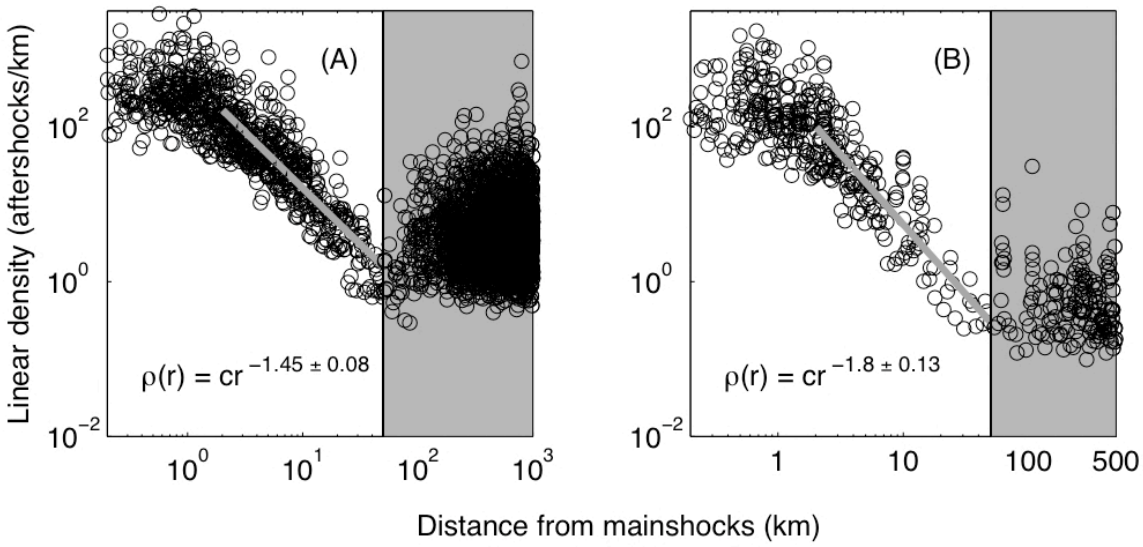
assigning weights by the number of days given in each panel. The result is an average background rate of 1.3 earthquakes/day/km for the combined data set; divided by the number of mainshocks this gives a background rate of 0.0068 earthquakes/day/km ($M \geq 2$) in real time. This predicts that the composite figure in panel (D) should be best fit with $b = 1.3 \times 25 = 32.5$, which agrees well with the $b=33$ solved for in the inversion. The goodness of fit is again verified by a low correlation coefficient for the data residuals ($r = 0.03$) and a Kolmogorov-Smirnoff test, both of which indicate that the null hypothesis that the underlying aftershock decay rate remains constant with time cannot be disproved at 95% and 65% confidence. The Kolmogorov-Smirnoff test is done by comparing the data to Monte Carlo simulated data drawn from the combination of a pure inverse law and a constant background distribution.

Most importantly, our data shows no evidence that the near field aftershocks transition to a steeper decay rate with time, as might be expected if there were a temporal transition from dynamic to static stress change triggering.



Supplementary Figure 5

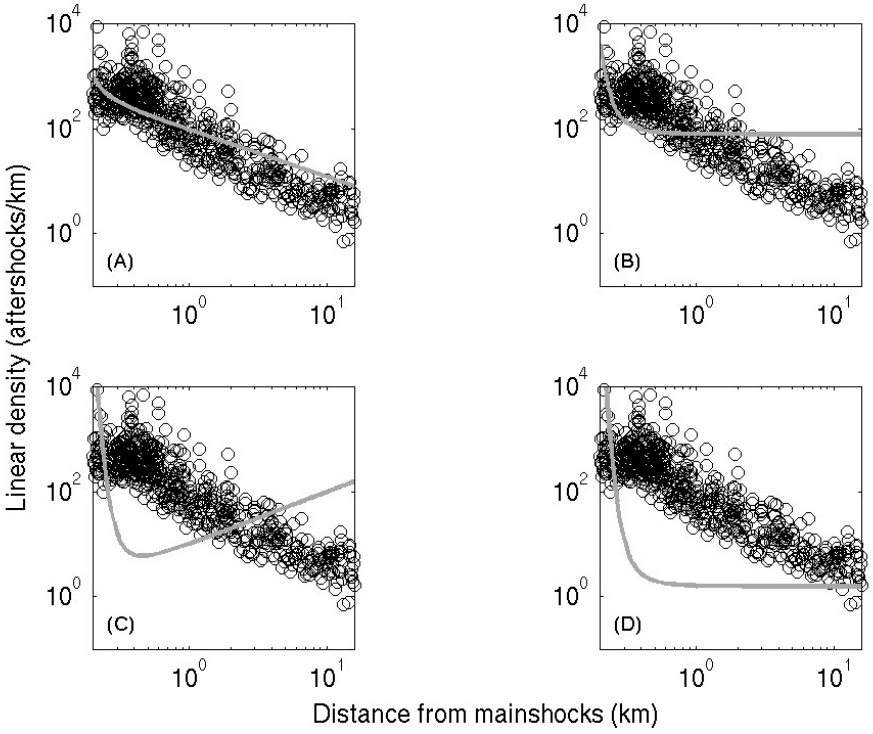
Here we provide the M 5—6 aftershock data given in Figure 3 over the fullest distance range possible (0 – 500 km). This data represents 9 mainshocks and 117 M≥3 aftershocks that occured over the first 2 days of each aftershock sequence. Aftershock density follows the same decay as it does for the M 2—4 mainshocks. Note that this data shows much less interference from background seismicity at long distances than the data in Supplementary Figure 1. This is for two reasons. First, the main data set (black dots) in Supplementary Figure 1 (B) uses 2355 mainshocks, with 5 minutes sampled after each mainshock, and goes down to M 2. This means that a total of 8.2 days worth of M ≥ 2 background seismicity is stacked into the plot. The present plot uses 9 mainshocks with 2 days of M ≥ 3 earthquakes sampled after each mainshock; thus it stacks 18 days of M 3 background seismicity or the equivalent of 1.8 days of M 2 background seismicity, about 20% of the equivalent number of background days used in Supplementary Figure 1. Second, the mainshocks for this plot were chosen from a time of overall quieter background activity. Because we search for aftershocks for a full two days after each mainshock, we did not use any mainshock that was preceded by a larger earthquake within 30 days. For Supplementary Figure 1, on the other hand, larger mainshocks were only precluded for 3 days (see the Methods section).



Supplementary Figure 6

Distance from the mainshock vs. aftershock linear density for unrelocated catalog data from Japan (JMA catalog, 1997-2003) excluding earthquakes deeper than 30 km (A) and Northern California (B), excluding Long Valley caldera (ANSS catalog, 1984-2004). Distances are measured between hypocenters. Aftershocks used in Northern California are M>2, in Japan M>2.5. Aftershocks occur within the first 30 minutes of their respective mainshocks. (We use the first 30 minutes rather than the first 5 minutes to obtain a reasonable amount of data for these sequences). M 3-4 mainshocks are used. Like the 30 minute aftershock data in Supplementary Figure 3, we can fit the densities with inverse power laws (gray lines). Based on a sample of reported catalog errors, the average 3D location error (98% confidence) is about 2.5 km for Northern California and 2.2 km for Japan. Thus the shortest distance fit is 2 km for both data sets. The fits extend to 50 km. The Japan data set uses 7305 mainshocks and 366 aftershocks; the Northern California data set 3014 mainshocks and 369 aftershocks. Beyond 50 km (the shaded region) the data is strongly contaminated with aftershocks of larger earthquakes and other sources of background seismicity (see Supplementary Figure 1 for further discussion), so the data can no longer be considered to be aftershocks of our target mainshock and are not fit.

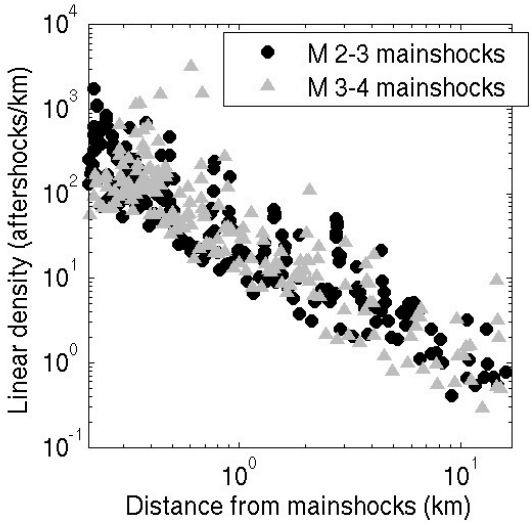
While the slope of the Japanese data is consistent with the slope we find in Southern California, the decay in Northern California is slightly steeper. Fitting the trend of background seismicity in Northern California using the same technique as in Figure 4 (main text) we find that the fault dimension, *D,* is about 0.3 smaller than in Southern California, indicating less complex fault structure. Assuming the same aftershock triggering mechanism in both locations, Equation 5 predicts that the exponent of the aftershock density decay in Northern California should be -1.7, which is within our measurement error bars.



Supplementary Figure 7

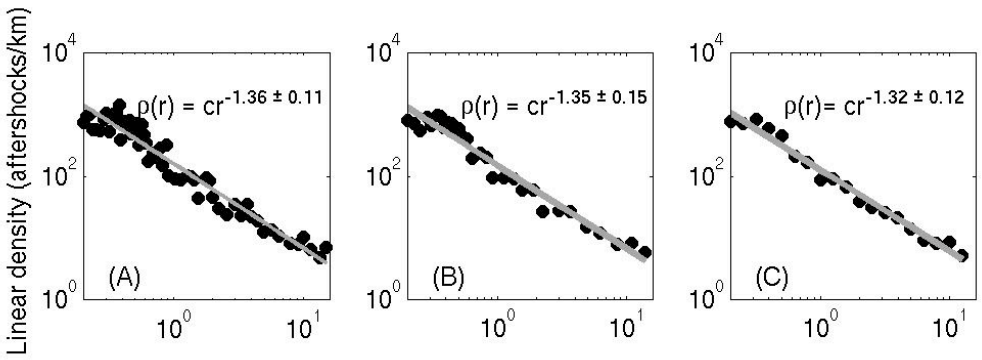
Predictions for variation of aftershock density with distance by the static stress change plus rate and state friction aftershock triggering model with various parameters (grey lines), plotted against the 30 minute aftershocks of M 3—4 aftershocks from Supplementary Figure 3 (black circles). All of the rate and state curves are associated with summed least squared errors which are higher than the error for a simple power law fit and the data residuals for all of the curves have high correlation coefficients, indicating poor fit. Density is aggregated for the entire dataset so the 30 minute background seismicity is equivalent to the number of mainshocks used (2355) multiplied by the 30 minute window, i.e., 49 days.

(A) Best fitting rate and state curve when a wide range of parameters are tried (see Methods). Best fit parameters are $D = 0.1$, background rate (c_1) = 0.0039 earthquakes/km/30 minutes or 0.19 M≥2 earthquakes/km/day, and $c_2/M_0 4\pi$, or $A\sigma$, = 0.33. The negative slope at long distances is due to the very low value of D. (Any value of D other than 1 will produce a trend in linear density in the the farfield.) For a good fit to 601 data points at 98% confidence, the correlation coefficient of the data residuals, r, should be less than 0.1. For the fit in (A), $r = 0.45$. (For a pure inverse power law fit r = 0.0079). Since D=0.1 is physically extreme and has not been observed, we also look for best fit solutions with set values of D. (B) The best fit with $D = 1$. Background rate = 1.6 earthquakes day/km, $A\sigma = 0.83$. $r = 0.71$. (C) The best fit with $D = 2$. Background rate = 3.2 earthquakes/day/km and $A\sigma = 1.7$. $r = 0.59$. (D) The best fit with $D = 1$ and the background rate constrained to the estimated real earthquake rate at 16 km distance of 0.03 earthquakes/day/km. The best fit solution gives $A\sigma = 1.9$, $r = 0.12$, and a summed least squared error that is 23 times higher than the pure inverse power law fit.



Supplementary Figure 8

The distance from the mainshock hypocenter vs. aftershock linear density for 200 aftershocks of M 2-3 mainshocks (black circles) and 200 aftershocks of M 3-4 mainshocks (gray triangles). Aftershocks are M > 2 aftershocks from the first 30 minutes of each aftershock sequence. When the same total number of aftershocks is used from each data set, the data points lie on top of each other. This indicates that the distribution of aftershock distances is independent of mainshock magnitude. Aftershock sequences of individual small mainshocks may appear compact simply because they have few aftershocks and the density decreases with distance, but there is in fact no correlation between the distance at which an aftershock occurs and the magnitude of its mainshock (correlation coefficient=-0.017, 290 data points between 0.2 and 16 km).



Supplementary Figure 9

Demonstration that our results are not highly sensitive to our choices of density measurement method and smoothing parameter. We use the first 30 minutes of data from M 3-4 mainshocks, over the distance range of 0.2 - 16 km. The densities in (A) and (B) are measured with the nearest neighbor method, with k = 10 (A) and k = 20 (B), meaning that the densities are estimated by taking the inverses of the lengths of the boxes needed to contain groups of 10 and 20 neighboring earthquakes, respectively. Densities in (C) are measured by counting the number of earthquakes in bins of equal length on a logarithmic scale (e.g. from making a histogram on a logarithmic scale). The slopes in these three plots are the same as in figures in the main text, where we measured density using the nearest neighbor method with k = 1. An inverse power law remains an excellent fit to the data, and the best fit power law exponents are consistent with the value of n found in the main text. The primary difference is that the data is more strongly smoothed. This allows for better visual presentation, but preserves less information, leading to a somewhat less well constrained fit for the inverse power law exponent.